

Quality assessment of digital X-ray chest images using an anthropomorphic chest phantom

A V Vodovatov¹, I G Kamishanskaya², A A Drozdov² and C Bernhardsson³

¹ Saint Petersburg Research Institute of Radiation Hygiene n. a. Professor P V Ramzaev, 197101, Saint Petersburg, Russia

² Saint Petersburg State University, Faculty of Medicine, Saint Petersburg, Russia

³ Medical Radiation Physics, Department of Translational Medicine, Malmö, Lund University, Skåne University Hospital, 205 02 Malmö, Sweden

E-mail: vodovatoff@gmail.com

Abstract. The current study is focused on determining the optimal tube voltage for the conventional X-ray digital chest screening examinations, using a visual grading analysis method. Chest images of an anthropomorphic phantom were acquired in posterior-anterior projection on four digital X-ray units with different detector types. X-ray images obtained with an anthropomorphic phantom were accepted by the radiologists as corresponding to a normal human anatomy, hence allowing using phantoms in image quality trials without limitations.

1. Introduction

Accurate measures of both clinical image quality and assessment of patient radiation risk are essential for successful optimization of medical imaging with ionizing radiation. Optimization in diagnostic radiology is equivalent of finding the image acquisition technique that maximizes the perceived information content and minimizes the radiation risk or keeps it at a reasonably low level [1]. All the changes made to the examination protocols should be applied only after it is confirmed that the revised exposure settings do not degrade diagnostic features of the image.

Currently, there are a variety of different techniques for the assessment/measurement of image quality [2–5]. Objective methods, based on the description of the physical characteristics of the image and their relation to the exposure, are usually performed with test-objects and simple physical phantoms. Subjective (or observer performance) methods, based on the visual perception of the whole image or parts of it, are usually performed with anthropomorphic phantoms for medical imaging or with real patient data [2, 5]. However, results obtained using the objective methods are more accurate and reproducible; they allow both measuring the primary physical characteristics of the imaging system or the overall system performance [3, 4]. Unfortunately, establishing a direct relation between those physical image characteristics and clinical image quality is difficult [3, 5]. On the other hand, the observer performance methods allow evaluating the quality of the whole imaging chain, thus measuring the image quality. Although there is a risk that these methods are influenced by a subjective nature of the observer, they consider the professional opinions and demands of radiologists [2, 5].

Observer performance methods are either based on the detection of the lesions (Receiver Operating Characteristics, ROC, and its variants) [4], or on the assessment of the visibility of the anatomical structures (visual grading analysis, VGA, or Image criteria, IC), the latter being more easily



implemented in the clinical practice [5–7]. Current set of image criteria for conventional radiography of adult patients have been presented by the European Commission [8]. These quality criteria have however not been updated since they were presented in 1996 and were primarily designed for the analogue X-ray equipment. Russian national radiological guidelines [9] suggest performing chest imaging in frontal projection on 80–100 kV, and the majority of the Russian X-ray unit vendors comply with those standards. Using relatively low tube voltage compared to the European and International practices (120–140+ kV) leads to fewer possibilities of lesion detection and yields to a higher exposure of the patient. Additionally, low tube voltage requires relatively high tube current-time product, resulting in a significantly higher exposure of the patient.

Hence, the main aim of the current study was to evaluate the clinical quality of chest images acquired at different tube voltages, and to select the optimal one based on the radiologist's opinion. The second aim of the study was to select the proper image quality assessment methodology that would be flexible and could be used as a tool for different optimization procedures based on end users (radiologists) subjective image quality assessment.

2. Materials and methods

The initial image quality criteria's were taken from Tingberg et al. [2], Uffmann et al. [7] and the European Commission recommendations [8]. Additional criteria were added by radiologists from the participating hospitals with more than 10 year working experience with digital X-ray units. Twelve criteria were selected, corresponding to the different anatomical regions of the chest. Complex anatomical regions (e.g. visibility of the vascular pattern in the peripheral areas of the lungs superimposed on ribs) were divided into the sub-regions.

The use of an anthropomorphic phantom instead of real patients allowed investigating the X-ray tube parameters otherwise impossible. Hence, in the present study it was possible to perform investigations of a variety of different protocols using the same object. The study was conducted using the Multipurpose Chest Phantom N1 "Lungman" (Kyoto Kagaku Co., Japan) anthropomorphic phantom [10, 11], simulating the anatomy of the chest of a healthy male (174 cm height, 75 kg weight). No pathologies (e.g. nodules, lesions, etc.) were added. During the image acquisition, the phantom was positioned in a standard posterior-anterior (PA) position for a patient undergoing chest examination according to a typical clinical protocol [5, 9]. Four digital X-ray units with different detector types that are commonly used for chest screening examinations (all manufactured by CSJC "Electron", Russia) commonly used for chest screening examinations in the radiology department of the Mariinsky hospital were selected for the study. Chest images were acquired with an automated exposure control at 3 different tube voltages: standard for Russia (90 kV), lowest used in clinical practice (60 kV) and highest used in clinical practice (120 kV). As a result, four sets (one set for every X-ray unit) of three images were constructed. Dose-area product (DAP) was measured using DRK-1 DAP-meters (DOZA, Russia), integrated into the X-ray units. The DAP-meters were calibrated by the vendor prior to the study. Effective dose was calculated using PCXMC software (STUK, Finland), using tissue weighting factors from ICRP publication 103 [12].

The trial on diagnostic image quality assessment was conducted at the radiology departments of two major university hospitals in Saint Petersburg: State Mariinsky hospital and Military Medical Academy. Thirteen radiologists from both hospitals (ten from Mariinsky hospital and three from Military Medical Academy) were selected as experts for the trial: six of them had less than 10 years of experience and seven of them had more than 10 years of experience, composing a representative group for the Russian radiological practice in the region. All the participants had a prior experience of working on the selected models of X-ray units. The phantom images were exported as DICOM files from the hospital PACS server. For each X-ray unit three images acquired at 60 kV (minimal tube voltage allowing chest imaging), 90 kV (median for the tube voltage range) and 120 kV (maximum tube voltage allowed by the generator capabilities) were selected. The images were processed by the X-ray units' software according to the standard settings for the chest protocol. Manual post-processing of the images (i.e. windowing, filtration) was disallowed due to the methodology of the trial

(segmentation and presentation of the resulting images). The regions and sub-regions of the images, corresponding to the selected criteria, were cut from the original images, shuffled and merged. This resulted in 25 separate triplets of images for each X-ray unit (figure 1). The resulting images were presented to the observers on a dedicated workstation with a diagnostic monitor (NEC MD21GS-2MP-BB, 21 inch, 3000×3000 pixel resolution), calibrated by the vendor.

Preliminary tests using the IC assessment method indicated that the digital images acquired on the X-ray units participating in the trial complied with the proposed criteria regardless of the parameters of image acquisition (tube voltage, specifically). All the images were marked as acceptable or above average quality. The current study was performed using a combination of the VGA and IC methods, thus allowing comparing the visibility of the suggested image criteria between the images acquired on different tube voltages. For each participating X-ray unit, images acquired on different tube voltages with the AEC were extracted from the PACS. After that, the areas corresponding to the relevant anatomic regions and sub-regions were cut from the original digital images. Further, for each region and sub-region cut, areas were shuffled to form a random sequence and merged together into one resulting image (triplet) for evaluation. An example of such resulting image is presented in figure 1.

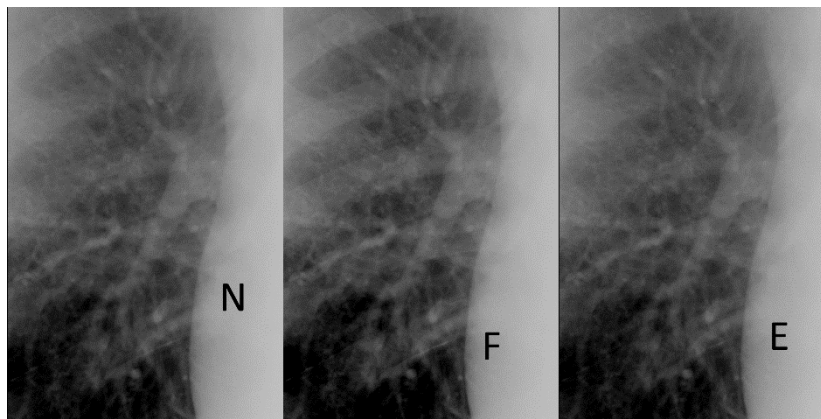


Figure 1. Example of the resulting image for the criterion 3 (Hilum of the lung) for the FC X-ray unit. Segments corresponding to the anatomical region of the hilum were cut from the chest images acquired on a different tube voltage (60, 90 and 120 kV), shuffled and merged. The letters (N, F, E) encode the tube voltage.

Instead of assigning a score based evaluation system on the quality of the criteria visualisation, the observers were required to rank the presented anatomic regions (figure 1), assigning first, second and third places. In case of a tie, each image was assigned the closest highest score. Each image segment received a score according to an assigned place. The image score was calculated as the sum of all the points for the segments. The image with the least overall score was considered the best.

A guidance on the selected image quality assessment method was distributed among the observers. It included a list of the criteria with the corresponding anatomic regions marked on the images of the chest. A series of test trials were performed to assure that the selected observers became familiar with the proposed methodology of the image quality assessment. The image rankings were converted into scores and processed using Kruskal-Wallis non-parametric test, using Statistica 10 software. The dependent variables included the rating scores for the image criteria; tube voltage was selected as the independent variable. Differences were considered significant with a $p < 0.05$.

3. Results and discussion

The overall image scores per X-ray unit were obtained by summing the scores for the separate image criterion for each observer. The results are presented in figure. 2 for each of the selected tube voltages (60, 90 and 120 kV) and for each X-ray unit. The difference between different tube voltage was insignificant only for the FC X-ray unit (Kruskal-Wallis test $p = 0.6$). The 60 kV protocol was selected as the best for the ARC1 X-ray unit and the 120 kV protocol was selected as the poorest for the KRT X-ray unit. No statistically significant differences between the remaining protocols were observed for those units. For the ARC2 X-ray unit (the only unit with an indirect flat-panel detector) statistically significant differences were observed between all three tube voltages. Also for this unit, the 60 kV image was selected as the best and the image acquired with 120 kV as the worst. Mean observer scores

for the separate image criteria (with the image scores for the sub-regions averaged into the scores for the whole region) and the data on the significance of the differences are presented in figure 2. The effective dose was calculated based on the DAP for the reference adult [12].

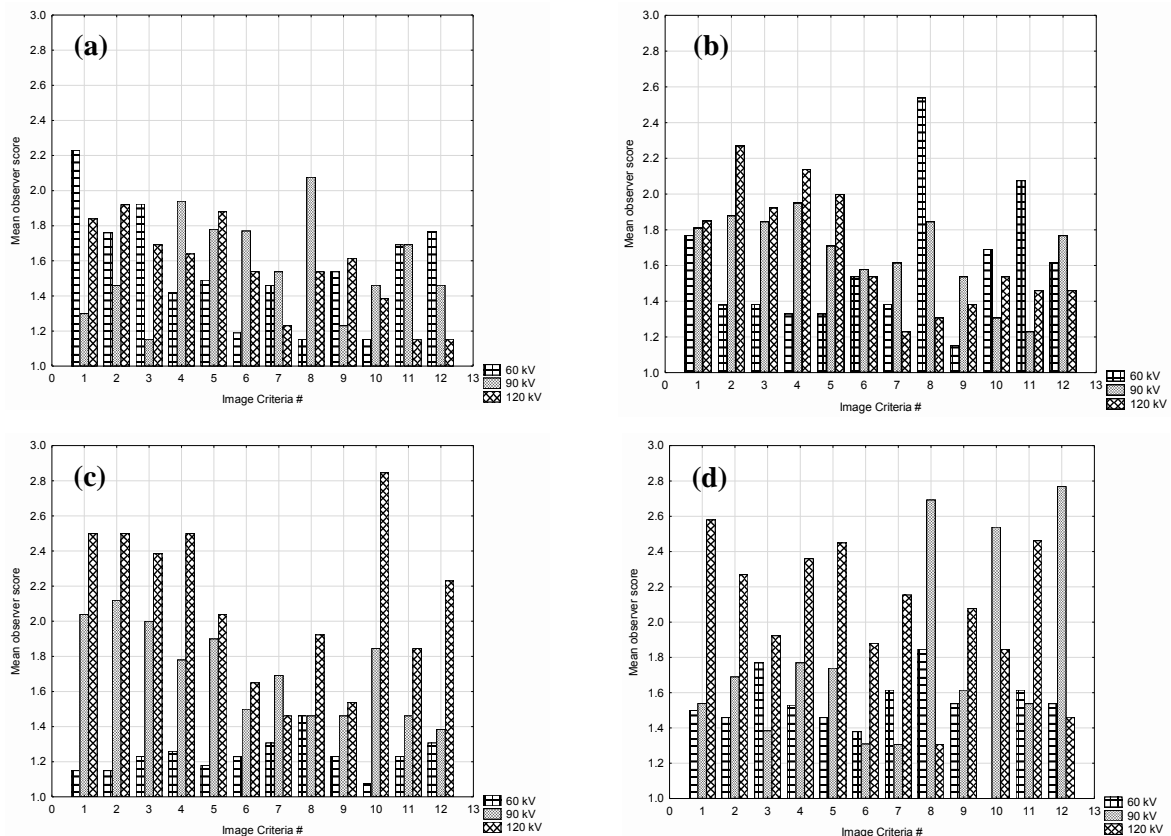


Figure 2. Mean observer scores for the selected image quality criteria for the FC (a), ARC1 (b), ARC2 (c) and KRT (d) X-ray units, respectively. Image scores for the sub-regions were averaged into the scores for the whole region. The best image quality corresponds to the lowest image score.

The Russian radiological practice lacks systematic approach to the image quality description and assessment, as the majority of the radiologists tend to evaluate the whole image basing their decision on the initial visual impression. Another factor is that most radiologists and operators work simultaneously with both analogue and digital X-ray units. Unsatisfactory digital images are considered to be acceptable because they still provide better diagnostic quality compared to images acquired on analogue units. Hence, the general demands on digital image quality are lower. In this case, using only the Image Criteria method would lead to the misinterpretation of the results. As mentioned before, in the preliminary trials based on the IC method, all images were marked as acceptable or above average, with no statistically significant differences between the combinations of the parameters of the examination (data not provided). Allowing the observers to rank the presented images by the visibility of the different anatomical structures (relating them to the different image quality criteria) tend to be the most feasible choice [5, 13].

Segmentation of the images, presenting only the cropped parts, corresponding to a specific anatomic region is a forced measure needed to perform the present study. Although it complicates the preparation of the trial, and disallows manual post-processing of the images, it is currently the only solution to force the observers to focus on the specific regions of the image, thus making the result reliable [2]. The use of the anthropomorphic phantoms in the trials was accepted by the radiologists, as long as the image quality assessment was focused on the visibility of the normal anatomy. Using the phantoms allowed also comparing different imaging protocols and X-ray units without any limitations

due to the ethical considerations [10]. Nevertheless, the set of the chest image quality criteria used in the present study was originally developed for real patients.

It is generally accepted, that the diagnostic image quality decreases with the increase of the mean energy of the X-ray beam due to the significant reduction of image contrast [5, 7]. The results of the current trial confirm this. For three out of four X-ray units and for the majority of the observers, 60 kV and 90 kV images had the lowest, and 120 kV images – the highest mean image scores (figure 1). For the FC X-ray unit the difference between three imaging protocols were not statistically significant. It is worth mentioning that the clear differences between the protocols, hence indicating the consensus of the observers, were visible only for the ARC2 X-ray unit with a flat-panel detector. The 60 kV and 90 kV or 120 kV protocols were similar for the X-ray units using a CCD-matrix detector.

It was not possible to determine the statistically significant differences for the majority of the separate image criteria, although they follow the same trend. It is visible from figure 2 that 60 kV images were mainly marked as best and 120 kV images – as worst, for the majority of the image criteria. According to our expectations the image criteria's were subdivided into three groups: best visible on low tube voltage; best visible on high tube voltage and criteria not affected by tube voltage at all. That grouping was based on the anatomical structure and the location of the relevant area. The anatomical structures located behind large and dense objects (i.e. thoracic aorta, located behind the sternum and the heart) were expected to be better visualized using high tube voltage. On the contrary, subtle and soft structures (pulmonary parenchyma) were expected to be better visualized using low tube voltage. It was expected, that the observers scores would be influenced by the differences in the visualization. According to the results (figure 2) low (60 kV and 90 kV) tube voltages were the best for all the three groups, contradictory to the results expected for group 2 and 3. Only the visualization of criteria 7 and 9 conformed to our expectations. It was not possible to identify any significant differences between the radiologists from both of the participating hospitals, or with different working experience. The diversity of the opinions of the observers can be explained by the personal preferences due to the different levels of education and training.

The preference of the low tube voltage for the majority of the criteria can be explained by different factors. Firstly, radiologists from the both hospitals commonly use 70–100 kV tube voltage range for the real patients, thus they intentionally or unintentionally preferred segments of images with similar appearance. Secondly, the radiologists interpreted a “good/best image” not as an “image with maximum information content”, or, in case of the current study, as an “image with the best representation of the normal anatomy of the chest” but as an “image with best appearance”. Despite the efforts to focus the observer's attention on the individual anatomic areas and particular anatomic details, it can be concluded that the majority of them still tend to evaluate the whole image presented. The decisions were made without considering the actual visibility of certain anatomical structures. This trend can be changed by additional training of the radiologists, introducing the concepts of image quality and quality criteria with an addition of the algorithmic approach to the image evaluation [13]. The visual appearance of the presented image segments acquired on 60 kV was more detailed in contrast compared with the others. That can be explained by the energy of the beam corresponding to the optimal energy for the CCD and CsI based flat panel detectors, complemented by the unsatisfactory X-ray unit image processing. The results for the ARC2 X-ray unit correspond to the results described by Uffman et al. [7], with low tube voltage (90 kV) protocol being significantly better compared to 121 and 150 kV protocols for the flat-panel detector system.

In order to fully evaluate three different imaging protocols, the patient dose should be considered as well. DAP and effective dose for the 60 kV imaging protocol are higher by a factor of 1.5–2 and 2–2.5 compared to the 90 and 120 kV, protocols respectively. In this case, using high tube voltage is beneficial, considering that screening examinations are performed mostly on healthy patients. The most viable approach to manage patient dose is to use the low-dose (high voltage) protocols for the X-ray units with the CCD-matrix detector and try to lower the patient dose for the low tube voltage protocol by adjusting the AEC settings and upgrading the image processing software for the X-ray unit with a flat-panel detector. Another solution would be to adjust the exposure of the detector at the

selected values of tube voltages, keeping the effective dose to the patient on the same level, and perform an additional image quality assessment study. That would allow for the assessment of the tube voltage, without considering the dose of the patient.

4. Conclusions

The results of our trial on clinical image quality indicate the following. We propose a diagnostic image quality assessment method using an anthropomorphic phantom. This method is based on the combination of the image criteria and visual grading analysis. It allows to evaluate the quality of the whole image or the separate anatomic areas, and can be used for the comparison of the different imaging protocols between the different X-ray units. The comparison of the chest images of the anthropomorphic phantom indicates that images, acquired on low tube voltage (60 and 90 kV) were preferable compared to images acquired on high tube voltage (120 kV);

High variations in the scores indicates a significant difference in education and training of the observers, both in requirements to the diagnostic image quality and experience in similar trials. Using the imaging protocols preferred by the majority of the radiologists with the current exposure settings will lead to significant additional exposure of the patient (up to the factor of 2–2.5 compared to high tube voltage protocols). We propose to use the low-dose (high voltage) protocols for the X-ray units with CCD-matrix considering the additional exposure of the patient and high variation in the image quality assessment. Low tube voltage should be used with X-ray units utilizing flat-panel detectors after the attempts to lower the patient dose by adjusting the AEC settings and upgrading the image processing software.

Acknowledgments

The authors would like to thank the staff of the radiology departments of the Saint Petersburg State Mariinsky hospital and Saint Petersburg Military Medical Academy for the possibility to use the X-ray equipment and the participation in the study.

This publication is part of a research work at Lund University, according to Swedish Institute scholarship (00110/2015) for A V Vodovatov. This work is partly financially supported by the Swedish Radiation Safety Authority (SSM).

References

- [1] Wouter J H V, Lucia J M K and Jacob G 2009 *Eur. Radiol.* **72** 209–17
- [2] Tingberg A 2000 *Quantifying the Quality of Medical X-Ray Images: An Evaluation Based on Normal Anatomy for Lumbar Spine and Chest Radiography* (Malmö: Lund University)
- [3] Tapiovaara M 2005 *Rad. Prot. Dosim.* **1–3** 116–9
- [4] Chakraborty D P and Berbaum K S 2004 *Med. Phys.* **8** 2313–30
- [5] Image Quality in Chest Radiography 2003 *Journal of the ICRU* **3**
- [6] Lanhede B, Båth M, Kheddache S et al. 2002 *Br J Radiol* **75** 38–49
- [7] Uffmann M, Neitzel U, Prokop M, Kabalan N, Weber M, Herold C J, Schaefer-Prokop C 2005 *Radiology* **235** 642–50
- [8] Carmichael J H E 1996 *European Guidelines on Quality Criteria for Diagnostic Radiographic Images* (Luxembourg: EUR-OP)
- [9] Abduraimov A B 2013 *National guidelines on X-ray diagnostics and therapy* (Moscow: GEOTAR-Media)
- [10] DeWerd L A and Kissick M 2014 *The Phantoms of Medical and Health Physics. Biological and Medical Physics, Biomedical Engineering* (New York: Springer New York)
- [11] Chest phantom N1 user manual: http://www.kyotokagaku.com/products/detail03/pdf/ph-1_manual.pdf
- [12] Tapiovaara M and Siiskonen T 2008 *PCXMC. A Monte Carlo program for calculating patient doses in medical x-ray examinations* (Helsinki: STUK-A231)
- [13] Folio L R 2012 *Chest imaging. An algorithmic approach to learning* (New York: Springer)